(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 September 2001 (20.09.2001)

PCT

(10) International Publication Number
WO 01/69928 A2

(51) International Patent Classification⁷: H04N 7/173, 5/445

(21) International Application Number: PCT/US01/07933

(22) International Filing Date: 12 March 2001 (12.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/524.056 13 March 2000 (13.03.2000) US

(71) Applicant (*for all designated States except US*): **SRI INTERNATIONAL** [US/US]; 333 Ravenswood Avenue, Menlo Park, CA 94025 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **JULIA, Luc** [US/US]; 607 Menlo Avenue, Menlo Park, CA 94025 (US). **VOUTSAS, Dimitris** [GR/GR]; 14 M. Pyrza Street, Neoi Epivates, GR-57019 Thessaloniki (GR). **CHEYER, Adam** [US/US]; 757 Cereza Drive, Palo Alto, CA 94306 (US).

(74) Agent: **ZILKA, Kevin J.**; Carlton Fields, P.O. Box 721030, San Jose, CA 95172-1030 (US).

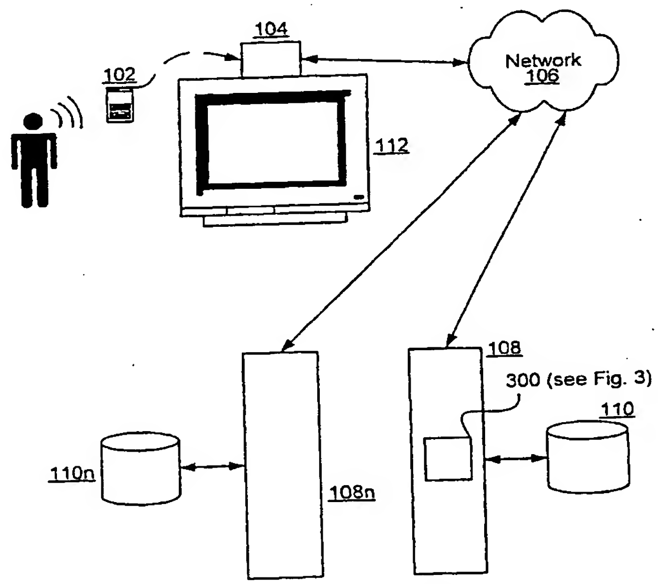
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM, METHOD AND ARTICLE OF MANUFACTURE FOR NAVIGATING NETWORK-BASED ELECTRONIC MULTIMEDIA CONTENT USING SPOKEN NATURAL LANGUAGE INPUT



(57) Abstract: A system, method, and article of manufacture are provided for navigating an electronic multimedia data source by means of spoken natural language. A spoken natural language input request is received from a user and interpreted. The resulting interpretation of the request is used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network multimedia data sources. The information is transmitted to a local user device.

WO 01/69928 A2

applies equally to surfing the Internet/Web or other networks for general information, multimedia content, or e-commerce transactions.

In general, the existing navigational systems for browsing electronic databases and data warehouses (search engines, menus, etc.), have been designed without navigation via spoken natural language as a specific goal. So today's world is full of existing electronic data navigation systems that do not assume browsing via natural spoken commands, but rather assume text and mouse-click inputs (or in the case of TV remote controls, even less). Simply recognizing voice commands within an extremely limited vocabulary and grammar -- the spoken equivalent of button/click input (e.g., speaking "channel 5" selects TV channel 5) -- is really not sufficient by itself to satisfy the objectives described above. In order to deliver a true "win" for users, the voice-driven front-end must accept spoken natural language input in a manner that is intuitive to users. For example, the front-end should not require learning a highly specialized command language or format. More fundamentally, the front-end must allow users to speak directly in terms of what the user ultimately wants -- e.g., "I'd like to see a Western film directed by Clint Eastwood" -- as opposed to speaking in terms of arbitrary navigation structures (e.g., hierarchical layers of menus, commands, etc.) that are essentially artifacts reflecting constraints of the pre-existing text/click navigation system. At the same time, the front-end must recognize and accommodate the reality that a stream of naïve spoken natural language input will, over time, typically present a variety of errors and/or ambiguities: e.g., garbled/unrecognized words (did the user say "Eastwood" or "Easter"?) and under-constrained requests ("Show me the Clint Eastwood movie"). An approach is needed for handling and resolving such errors and ambiguities in a rapid, user-friendly, non-frustrating manner.

What is needed is a methodology and apparatus for rapidly constructing a voice-driven front-end atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the step-by-step browsing architecture of the existing navigation system, and wherein any errors or ambiguities in user input are rapidly and conveniently resolved. The solution to this need should be compatible with the constraints of a multi-user, distributed environment such as the Internet/Web or a proprietary high-bandwidth

transmitted from device 202 to requests processing logic 300, hosted on a local speech processor, for processing and interpretation. In the preferred embodiment illustrated in Figure 1b, the local speech processor is conveniently integrated as part of communications box 104, although implementation in a physically separate (but communicatively coupled) unit is also possible as will be readily apparent to those of skill in the art. The voice data is processed by the components of request processing logic 300 in order to understand the user's request and construct an appropriate query or request for navigation of remote data source 110, in accordance with the interpretation process exemplified in Figures 4 and 5 as discussed in greater detail below.

The resulting navigational query is then transmitted electronically across network 106 to data source 110, which preferably resides on a central server or servers 108. As in Figure 1a, data source 110 may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source 110 are then navigated -- i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user -- preferably using the process of Figures 4 and 5 as described in greater detail below. Once the desired information has been retrieved from data source 110, it is electronically transmitted via network 106 to the user for viewing on client display device 112.

In one embodiment in accordance with Figure 1b and well-suited for the home entertainment setting, voice input device 102 is a portable remote control device with an integrated microphone, and the voice data is transmitted from device 102 preferably via infrared (or other wireless) link to the local speech processor. The local speech processor is coupled to communications network 106, and also preferably to client display device 112 (especially for purposes of query refinement transmissions, as discussed below in connection with Figure 4, step 412), and preferably may be integrated within or coupled to communications box 104. In addition, especially for purposes of a home entertainment application, display device 112 is preferably a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such

Practitioners will further appreciate, in light of the above teachings, that if mobile information appliance 202 is equipped with sufficient computational processing power, then a mobile variation of the client-side architecture exemplified in Figure 2 may similarly be implemented. In that case, the modules corresponding to request processing logic 300 would be embodied locally in the computational resources of mobile information appliance 202, and the logical flow of data would otherwise follow in a manner analogous to that previously described in connection with Figure 1b.

As illustrated in Figure 2, multiple users, each having their own client input device, may issue requests, simultaneously or otherwise, for navigation of data source 210. This is equally true (though not explicitly drawn) for the embodiments depicted in Figures 1a and 1b. Data source 210 (or 100), being a network-accessible information resource, has typically already been constructed to support access requests from simultaneous multiple network users, as known by practitioners of ordinary skill in the art. In the case of server-side speech processing, as exemplified in Figures 1a and 2, the interpretation logic and error correction logic modules are also preferably designed and implemented to support queuing and multi-tasking of requests from multiple simultaneous network users, as will be appreciated by those of skill in the art.

It will be apparent to those skilled in the art that additional implementations, permutations and combinations of the embodiments set forth in Figures 1a, 1b, and 2 may be created without straying from the scope and spirit of the present invention. For example, practitioners will understand, in light of the above teachings and design considerations, that it is possible to divide and allocate the functional components of request processing logic 300 between client and server. For example, speech recognition -- in entirety, or perhaps just early stages such as feature extraction -- might be performed locally on the client end, perhaps to reduce bandwidth requirements, while natural language parsing and other necessary processing might be performed upstream on the server end, so that more extensive computational power need not be distributed locally to each client. In that case, corresponding portions of request processing logic 300, such as speech recognition engine 310 or portions

applications. As one other example, IBM offers the ViaVoice speech recognition engine, including a low-cost shrink-wrapped version available through popular consumer distribution channels. Basically, a speech recognition engine processes acoustic voice data and attempts to generate a text stream of recognized words.

Typically, the speech recognition engine is provided with a vocabulary lexicon of likely words or phrases that the recognition engine can match against its analysis of acoustical signals, for purposes of a given application. Preferably, the lexicon is dynamically adjusted to reflect the current user context, as established by the preceding user inputs. For example, if a user is engaged in a dialogue with the system about movie selection, the recognition engine's vocabulary may preferably be adjusted to favor relevant words and phrases, such as a stored list of proper names for popular movie actors and directors, etc. Whereas if the current dialogue involves selection and viewing of a sports event, the engine's vocabulary might preferably be adjusted to favor a stored list of proper names for professional sports teams, etc. In addition, a speech recognition engine is provided with language models that help the engine predict the most likely interpretation of a given segment of acoustical voice data, in the current context of phonemes or words in which the segment appears. In addition, speech recognition engines often echo to the user, in more or less real-time, a transcription of the engine's best guess at what the user has said, giving the user an opportunity to confirm or reject.

In a further aspect of step 404, natural language interpreter (or parser) 320 linguistically parses and interprets the textual output of the speech recognition engine. In a preferred embodiment of the present invention, the natural-language interpreter attempts to determine both the meaning of spoken words (semantic processing) as well as the grammar of the statement (syntactic processing), such as the Gemini Natural Language Understanding System developed by SRI International. The Gemini system is described in detail in publications entitled "Gemini: A Natural Language System for Spoken-Language Understanding" and "Interleaving Syntax and Semantics in an Efficient Bottom-Up Parser," both of which are currently available online at http://www.ai.sri.com/natural-language/projects/arpa-sls/nat-lang.html. (Copies of those publications are also included in an information disclosure statement submitted herewith, and are incorporated herein by this reference). Briefly, Gemini

search techniques. For some applications, an embodiment of the present invention may be implemented in which only access to a particular data source (such as a particular vendor's proprietary content database) is supported; in that case, step 405 may be trivial or may be eliminated entirely.

Step 406 attempts to construct a navigation query, reflecting the interpretation of step 404. This operation is preferably performed by query construction logic 330.

A "navigation query" means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information. In other words, a navigation query is constructed such that it includes whatever content and structure is required in order to access desired information electronically from a particular database or data source of interest.

For example, for many existing electronic databases, a navigation query can be embodied using a formal database query language such as Standard Query Language (SQL). For many databases, a navigation query can be constructed through a more user-friendly interactive front-end, such as a series of menus and/or interactive forms to be selected or filled in. SQL is a standard interactive and programming language for getting information from and updating a database. SQL is both an ANSI and an ISO standard. As is well known to practitioners, a Relational Database Management System (RDBMS), such as Microsoft's Access, Oracle's Oracle7, and Computer Associates' CA-OpenIngres, allow programmers to create, update, and administer a relational database. Practitioners of ordinary skill in the art will be thoroughly familiar with the notion of database navigation through structured query, and will be readily able to appreciate and utilize the existing data structures and navigational mechanisms for a given database, or to create such structures and mechanisms where desired.

In accordance with the present invention, the query constructed in step 406 must reflect the user's request as interpreted by the speech recognition engine and the NL parser in step 404. In embodiments of the present invention wherein data source 110 (or 210 in the corresponding embodiment of Figure 2) is a structured relational database or the like, step 406 of the present invention may entail constructing an

interpreted language that has built-in support for common web protocols like HTTP and FTP, and popular data types like HTML and XML. WebL's implementation language is Java, and the complete source code is available from Compaq. In addition, step 520 is preferably performed dynamically when necessary -- in other words, on-the-fly in response to a particular user query -- but in some applications it may be possible to scrape relatively stable (unchanging) web sites of likely interest in advance and to cache the resulting template information.

It will be apparent, in light of the above teachings, that preferred embodiments of the present invention can provide a spoken natural language interface atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the linear browsing architecture or other artifacts of an existing menu/text/click navigation system. For example, users of an appropriate embodiment of the present invention for a video-on-demand application can directly speak the natural request: "Show me the movie 'Unforgiven'" -- instead of walking step-by-step through a typically linear sequence of genre/title/actor/director menus, scrolling and selecting from potentially long lists on each menu, or instead of being forced to use an alphanumeric keyboard that cannot be as comfortable to hold or use as a lightweight remote control. Similarly, users of an appropriate embodiment of the present invention for a web-surfing application in accordance with the process shown in Figure 5 can directly speak the natural request: "Show me a one-month price chart for Microsoft stock" -- instead of potentially having to navigate to an appropriate web site, search for the right ticker symbol, enter/select the symbol, and specify display of the desired one-month price chart, each of those steps potentially involving manual navigation and data entry to one or more different interaction screens. (Note that these examples are offered to illustrate some of the potential benefits offered by appropriate embodiments of the present invention, and not to limit the scope of the invention in any respect.)

c. Error Correction

Several problems can arise when attempting to perform searches based on spoken natural language input. As indicated at decision step 407 in the process of Figure 4, certain deficiencies may be identified during the process of query

remaining problems or deficiencies are identified at decision points 407 or 409. Further details and examples for this query refinement process are provided immediately below.

Consider again the example in which the user of a video-on-demand application wishes to see "Unforgiven" but can only recall that it was directed by and starred Clint Eastwood. First, it bears noting that using a prior art navigational interface, such as a conventional menu interface, will likely be relatively tedious in this case. The user can proceed through a sequence of menus, such as Genre (select "western"), Title (skip), Actor ("Clint Eastwood"), and Director ("Clint Eastwood"). In each case --especially for the last two items -- the user would typically scroll and select from fairly long lists in order to enter his or her desired name, or perhaps use a relatively couch-unfriendly keypad to manually type the actor's name twice.

Using a preferred embodiment of the present invention, the user instead speaks aloud, holding remote control microphone 102, "I want to see that movie starring and directed by Clint Eastwood. Can't remember the title." At step 402 the voice data is received. At step 404 the voice data is interpreted. At step 405 an appropriate online data source is selected (or perhaps the system is directly connected to a proprietary video-on-demand provider). At step 406 a query is automatically constructed by the query construction logic 330 specifying "Clint Eastwood" in both the actor and director fields. Step 407 detects no obvious problems, and so the query is electronically submitted and the data source is navigated at step 408, yielding a list of several records satisfying the query (e.g., "Unforgiven", "True Crime", "Absolute Power", etc.). Step 409 detects that additional user input is needed to further refine the query in order to select a particular film for viewing.

At that point, in step 412 query refinement logic 340 might preferably generate a display for client display device 112 showing the (relatively short) list of film titles that satisfy the user's stated constraints. The user can then preferably use a relatively convenient input modality, such as buttons on the remote control, to select the desired title from the menu. In a further preferred embodiment, the first title on the list is highlighted by default, so that the user can simply press an "OK" button to choose that selection. In a further preferred feature, the user can mix input modalities

desired city has now been clarified. Step 408 navigates the data source using that query and retrieves the desired weather information, which is then electronically transmitted in step 410 from network server 108 to client display device 112 via communications network 106.

It is worth noting that in some instances, there may be details that are not explicitly provided by the user, but that query construction logic 330 or query refinement logic 340 may preferably deduce on their own through reasonable assumptions, rather than requiring the use to provide explicit clarification. For example, in the example previously described regarding a request for a weather report, in some applications it might be preferable for the system to simply assume that the user means a weather report for his or her home area and to retrieve that information, if the cost of doing so is not significantly greater than the cost of asking the user to clarify the query. Making such an assumption might be even more strongly justified in a preferred embodiment, as described earlier, where user histories are tracked, and where such history indicates that a particular user or group of users typically expect local information when asking for a weather forecast. At any rate, in the event such an assumption is made, if the user actually intended to request the weather for a different city, the user would then need to ask his or her question again. It will be apparent to practitioners, in light of the above teachings, that the choice of whether to program query construction logic 330 and query refinement logic 340 to make make particular assumptions will typically involve trade-offs involving user conveience that can be assessed in the context of specific applications.

corresponding discussion of a "unified messaging" application. Another example is the InfoWiz interactive information kiosk developed by the assignee and described in the document entitled "InfoWiz: An Animated Voice Interactive Information System" available online at http://www.ai.sri.com/~oaa/applications.html. A copy of the InfoWhiz document is provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference. A further example is the "CommandTalk" application developed by the assignee for the U.S. military, as described online at http://www.ai.sri.com/~lesaf/commandtalk.html and in the following publications, copies of which are provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference:

- "CommandTalk: A Spoken-Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, DC, pp. 1-7, Association for Computational Linguistics

- "The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty-Seventh Annual Meeting of the ACL", pp. 183-190, University of Maryland, College Park, MD, Association for Computational Linguistics

- "Interpreting Language in Context in CommandTalk", 1999, by John Dowding and Elizabeth Owen Bratt and Sharon Goldwater, in "Communicative Agents: The Use of Natural Language in Embodied Systems", pp. 63-67, Association for Computing Machinery (ACM) Special Interest Group on Artificial Intelligence (SIGART), Seattle, WA

For some applications and systems, OAA can provide an advantageous platform for constructing embodiments of the present invention. For example, a representative application is now briefly presented, with reference to Figure 6. If the statement "show me movies starring John Wayne" is spoken into the voice input device, the voice data for this request will be sent by UI agent 650 to facilitator 600, which in turn will ask natural language (NL) agent 620 and speech recognition agent 610 to interpret the query and return the interpretation in *ICL* format. The resulting *ICL* goal expression is then routed by the facilitator to appropriate agents -- in this case, video-on-demand database agent 640 -- to execute the request. Video database agent 640 preferably includes or is coupled to an appropriate embodiment of query construction logic 330 and query refinement logic 340, and may also issue ICL

as including all such alterations, permutations, and equivalents as fall within the true spirit and scope of the present invention.

7. The method of claim 1, wherein at least a portion of a communications link between the portable microphone and the multimedia data source is the Internet.

8. The method of claim 1, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and a linguistic parser.

9. The method of claim 1, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

10. The method of claim 9, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

11. The method of claim 1, wherein the navigation query is constructed in the format of a database query language.

12. The method of claim 1, wherein at least one of the steps of rendering an interpretation and constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

13. The method of claim 12, wherein the computing device is a communications box.

14. The method of claim 13, wherein the portable microphone communicates wirelessly with the communications box.

15. The method of claim 1, wherein at least one of the steps of rendering an interpretation and constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

16. The method of claim 1, further comprising the step of soliciting additional input in response to one or more deficiencies encountered during the step of constructing a navigation query.

25. The method of claim 16, wherein the additional input is solicited by presenting a menu to the user on the client device of the user.

26. The method of claim 16, wherein the additional input is solicited by presenting a textual request for the additional input.

27. The method of claim 16, wherein the additional input is solicited by outputting an audible request for the additional input.

28. The method of claim 16, wherein the additional input is solicited by presenting a list of portions of the electronic data source that match the navigational query.

29. The method of claim 16, wherein the additional input received from the user is at least partially speech based.

30. The method of claim 16, wherein the additional input received from the user includes no spoken input.

32. The method of claim 1, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

33. The method of claim 1, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

34. The method of claim 1, wherein the selected portion of the multimedia data source provides access to at least one of audio and video programming.

35. A system for speech-based navigation of an electronic multimedia data source, the electronic multimedia data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken request for desired information from the user;

44. The system of claim 42, wherein the query construction logic constructs the query in the format of a database query language.

45. The system of claim 35, further comprising a computing device located locally with the user for hosting at least a portion of the language processing logic, and wherein the portable microphone is electronically coupled to the local computing device.

46. The system of claim 35, wherein at least a portion of the language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

47. The system of claim 46, wherein the computing device is a communications box.

48. The system of claim 46, wherein portable microphone communicates wirelessly with the communications box.

49. The system of claim 35, wherein at least a portion of the language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

50. The system of claim 35, further comprising user interaction logic for soliciting additional input in response to one or more deficiencies encountered during the constructing the navigation query.

51. The system of claim 50, wherein the deficiencies include unresolved words of the spoken request.

52. The system of claim 50, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

53. The system of claim 50, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of

(e) a code segment that that transmits the selected portion of the electronic multimedia data source from the network server to a primarily stationary, display device located locally with the user..

60. The computer program of claim 59, wherein the multimedia data source stores multimedia content including at least one of video content and audio content.

61. The computer program of claim 59, wherein the display device is a television.

62. The computer program of claim 61, wherein the selected portion of the multimedia data source stores information relating to television programming.

63. The computer program of claim 59, wherein the selected portion of the multimedia data source provides streaming video.

64. The computer program of claim 59, wherein at least a portion of a communications link between the portable microphone and the multimedia data source is wireless.

65. The computer program of claim 59, wherein at least a portion of a communications link between the portable microphone and the multimedia data source is the Internet.

66. The computer program of claim 59, further comprising a code segment that derives linguistic information by using a speech recognition engine and a linguistic parser.

67. The computer program of claim 59, further comprising a code segment that extracts an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

68. The computer program of claim 67, wherein code segment that extracts an input template dynamically scrapes the online scripted interface.

69. The computer program of claim 59, wherein the navigation query is constructed in the format of a database query language.

80. The computer program of claim 59, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.

81. The computer program of claim 59, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

82. The computer program of claim 59, wherein the selected portion of the multimedia data source provides access to at least one of audio and video programming.

83. The computer program of claim 59, further comprising:

(a) a code segment that receives the additional input from the user

(b) a code segment that refines the navigational query based on the additional input;

(c) a code segment that determines whether the navigational query is deemed adequate;

(d) a code segment that solicits additional input if the navigational query is not deemed adequate and repeating steps (a)-(c) until the navigational query is deemed adequate.

84. The computer program of claim 59, wherein a determination that deficiencies exist in the navigational query is at least partially based on a user-input statement that additional information is required.

85. The computer program of claim 59, wherein the additional input is solicited by presenting a menu to the user on the client device of the user.

86. The computer program of claim 59, wherein the additional input is solicited by presenting a textual request for the additional input.

87. The computer program of claim 59, wherein the additional input is solicited by outputting an audible request for the additional input.

95. The method of claim 91, wherein the selected portion of the multimedia data source provides streaming video.

96. The method of claim 91, wherein at least a portion of a communications link between the portable microphone and the multimedia data source is wireless.

97. The method of claim 91, wherein at least a portion of a communications link between the portable microphone and the multimedia data source is the Internet.

98. The method of claim 91, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and an NL parser.

99. The method of claim 91, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

100. The method of claim 99, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

101. The method of claim 91, wherein the navigation query is constructed in the format of a database query language.

102. The method of claim 91, wherein at least one of the steps of rendering an interpretation and constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

103. The method of claim 102, wherein the computing device is a communications box.

104. The method of claim 103, wherein the portable microphone communicates wirelessly with the communications box.

105. The method of claim 91, wherein at least one of the steps of rendering an interpretation and constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

(b) spoken language processing logic, operable to render an interpretation of the spoken natural language request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken natural language request;

(d) navigation logic, operable to select a portion of the electronic multimedia data source using the navigation query; and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic multimedia data source from the network server to a primarily stationary, display device located locally with the user.

116. The system of claim 115, wherein the electronic multimedia data source stores multimedia content including at least one of video content and audio content.

117. The system of claim 115, wherein the display device is a television.

118. The system of claim 117, wherein the selected portion of the multimedia data source stores information relating to television programming.

119. The system of claim 115, wherein the selected portion of the multimedia data source provides streaming video.

120. The system of claim 115, wherein the electronic communications infrastructure includes an Internet link.

121. The system of claim 115, wherein the spoken language processing logic includes speech recognition logic and an NL parsing logic for deriving linguistic information.

122. The system of claim 115, wherein the spoken language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

123. The system of claim 122, wherein the spoken language processing logic dynamically scrapes the online scripted interface.

the data source using the navigation query constructed by the query construction logic.

134. The system of claim 133, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

135. The system of claim 133, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

136. The system of claim 115, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

137. The system of claim 115, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

138. The system of claim 115, wherein the selected portion of the electronic multimedia data source provides access to at least one of audio and video programming.

139. A computer program embodied on a computer readable medium for utilizing spoken natural language for navigating an electronic multimedia data source, the electronic multimedia data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken natural language ("NL") request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken natural language request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that uses the refined navigation query to select a portion of the electronic multimedia data source; and

150. The computer program of claim 139, wherein at least one of the rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

151. The computer program of claim 150, wherein the computing device is a communications box.

152. The computer program of claim 151, wherein the portable microphone communicates wirelessly with the communications box.

153. The computer program of claim 139, wherein at least one of the rendering of the interpretation and the constructing of the navigation query are performed, at least in part, on a network computing device located remotely from the user.

154. The computer program of claim 139, wherein the code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the construction of the navigation query.

155. The computer program of claim 154, wherein the deficiencies include unresolved words of the spoken NL request.

156. The computer program of claim 154, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

157. The computer program of claim 154, wherein the code segment that solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query.

158. The computer program of claim 157, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

159. The computer program of claim 157, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

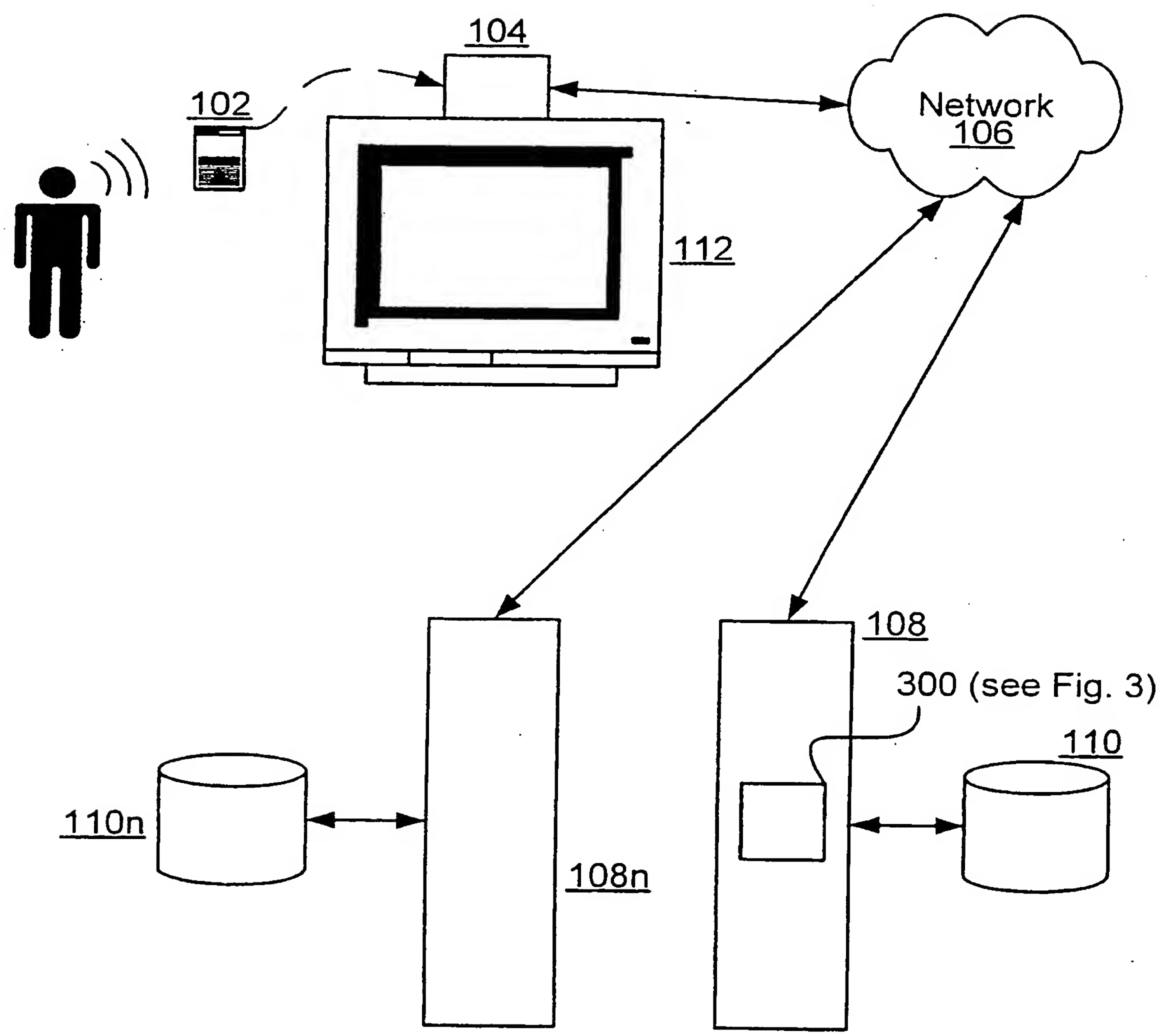


Fig. 1a

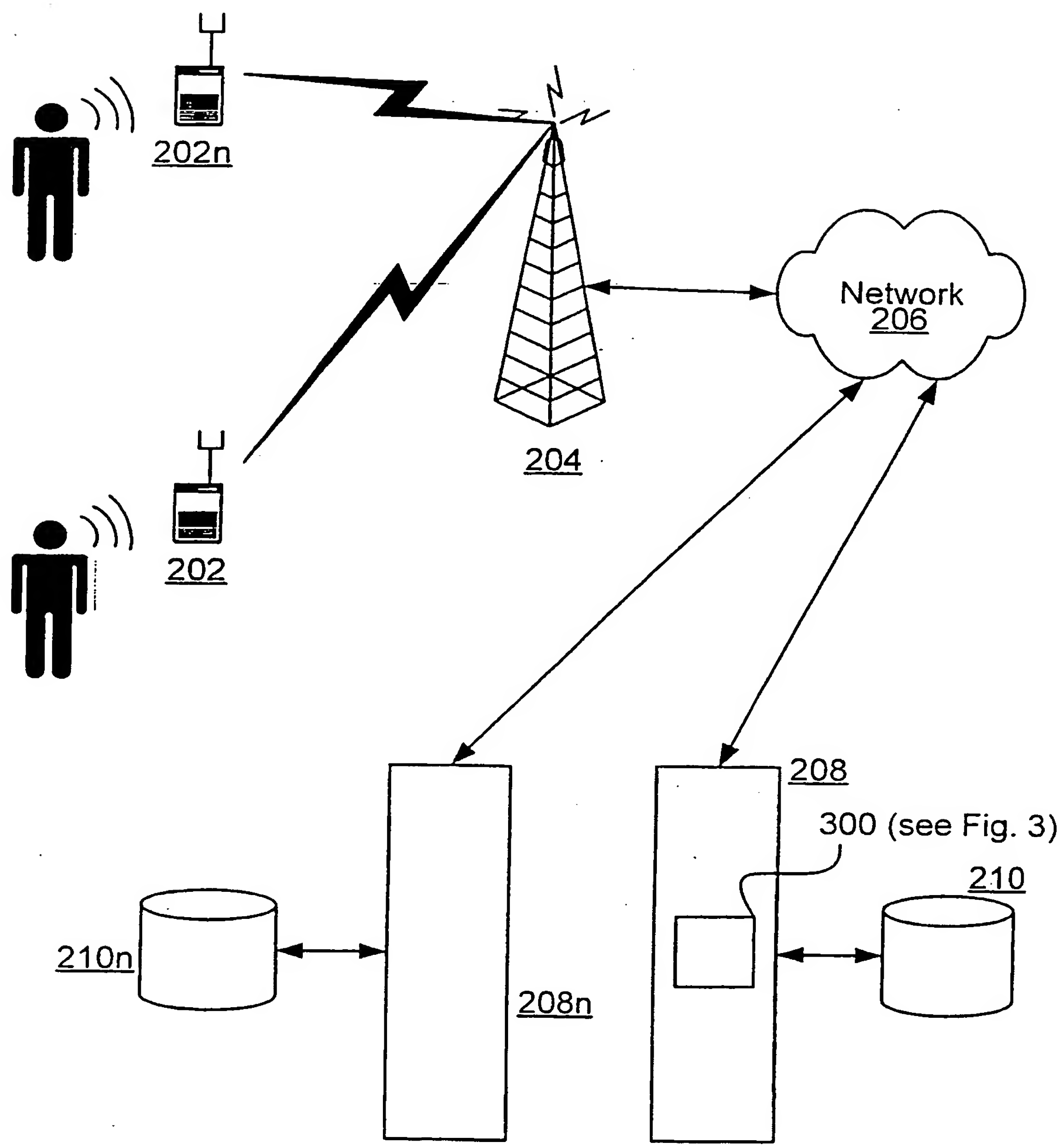


Fig. 2

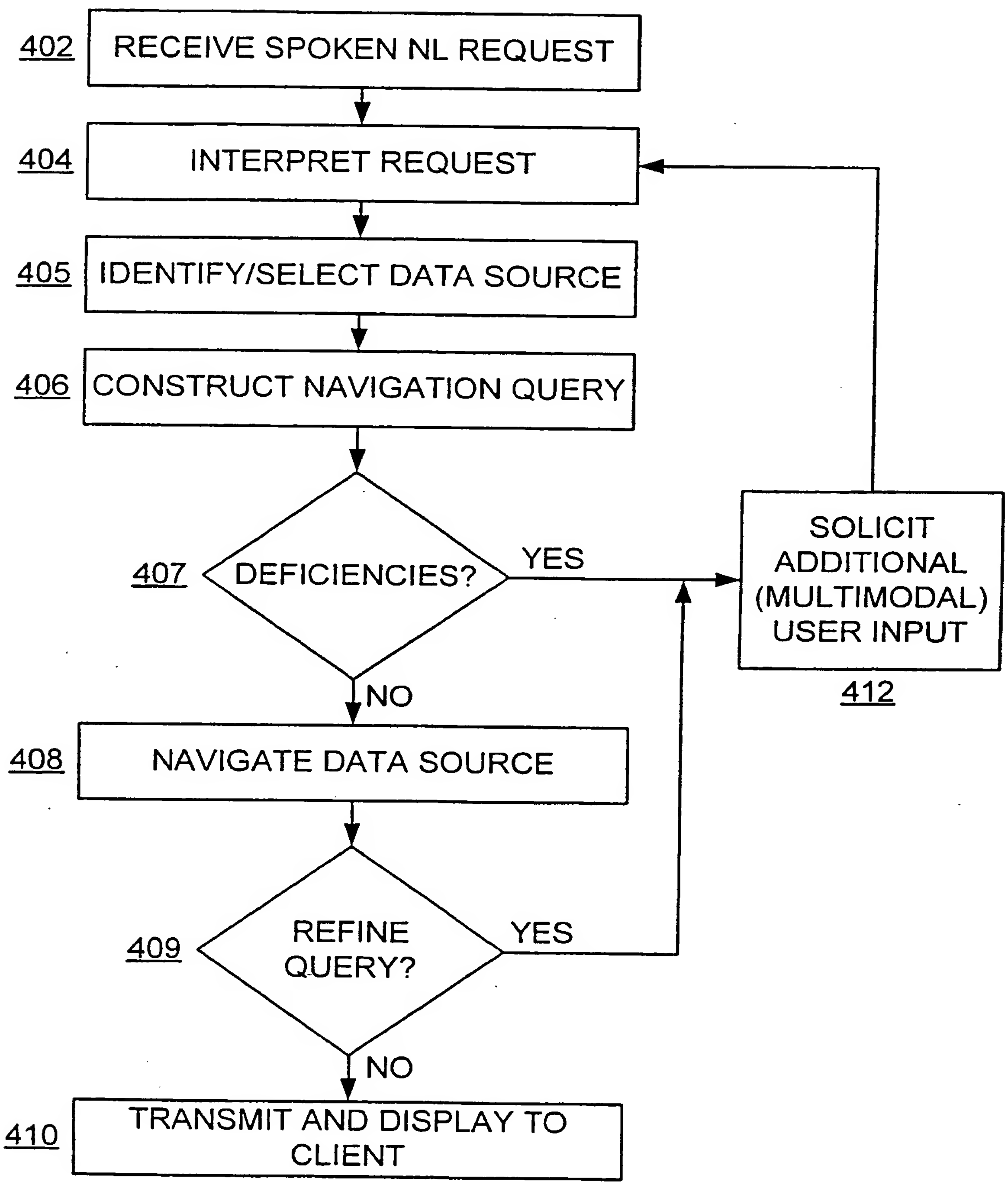

402
RECEIVE SPOKEN NL REQUEST
404
INTERPRET REQUEST
405
IDENTIFY/SELECT DATA SOURCE
406
CONSTRUCT NAVIGATION QUERY
407
DEFICIENCIES?
YES
NO
SOLICIT ADDITIONAL (MULTIMODAL) USER INPUT
412
408
NAVIGATE DATA SOURCE
409
REFINE QUERY?
YES
NO
410
TRANSMIT AND DISPLAY TO CLIENT

Fig. 4

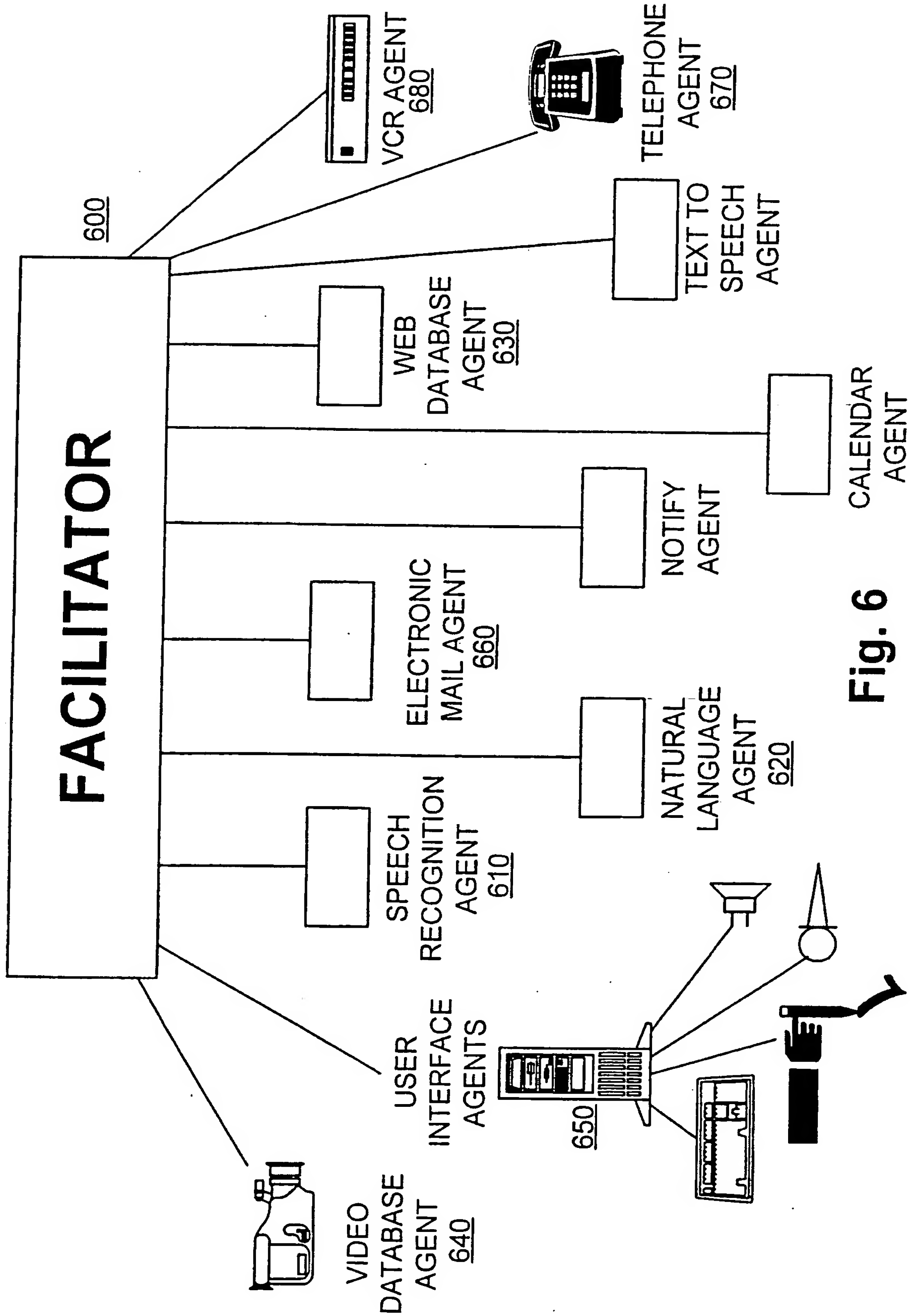

FACILITATOR
600
SPEECH RECOGNITION AGENT 610
NATURAL LANGUAGE AGENT 620
WEB DATABASE AGENT 630
VIDEO DATABASE AGENT 640
USER INTERFACE AGENTS
650
ELECTRONIC MAIL AGENT 660
TELEPHONE AGENT 670
VCR AGENT 680
TEXT TO SPEECH AGENT
NOTIFY AGENT
CALENDAR AGENT

Fig. 6